



兰州大学

# 毕业论文

(本科生)

论文题目（中文） 《数据结构》课程的学习报告

论文题目（英文） Study Report of Data Structures

学 生 姓 名 谭源

导师姓名、职称 蒙应杰

学生所属学院 信息科学与工程学院

专 业 计算机类

年 级 2019 级

兰州大学教务处

## 诚信责任书

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或在网上发表的论文。

本声明的法律责任由本人承担。

论文作者签名：\_\_\_\_\_ 日期：\_\_\_\_\_

## 关于毕业论文（设计）使用授权的声明

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文（设计）的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文（设计）的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文（设计）。本人离校后发表、使用毕业论文（设计）或与该毕业论文（设计）直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文（设计）研究内容：

- 可以公开
- 不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名：\_\_\_\_\_ 导师签名：\_\_\_\_\_

日期：\_\_\_\_\_ 日期：\_\_\_\_\_

# **《数据结构》课程的学习报告**

## **中文摘要**

在计算机科学中，数据结构是计算机中存储、组织数据的方式。这学期蒙老师先由数据结构的定义开始讲起，依次讲解了算法、线性表、栈和队列、串、数组和广义表、树形结构、图结构、排序、数据检索这几部分内容

不要仅仅把它当做广告，这里面有很多 latex 的用法说明

**关键词：** 兰朵儿， i 兰大易班， yuh

# **STUDY REPORT OF DATA STRUCTRES**

## **Abstract**

This essay explores the history of studies in analytical philosophy in China since the beginning of the last century, by dividing into three phases. It shows that, in these phases, analytic philosophy was always at a disadvantage in confronting serious challenges coming from both Chinese traditional philosophy and modern philosophical trends. The authors argue that Chinese philosophers have both done preliminary studies and offered their own analyses of various problems as well as some new applications of analytic philosophy especially in the latest period. Meanwhile, Chinese traditional philosophy was always trying to adjust its cultural mentality in the struggle with analytic philosophy, and accommodated in its own way the rationalistic spirit and scientific method represented in analytic philosophy.

**Key Words:** analytical philosophy; Chinese philosophers; philosophical analysis; dialogue in philosophy.

# 目 录

中文摘要 .....	I
英文摘要 .....	II
第一章 数据结构绪论 .....	1
1.1 数据结构的基本概念及研究内容 .....	1
1.2 数据结构的选择与评价 .....	2
第二章 算法 .....	3
2.1 算法的定义 .....	3
2.1.1 概述 .....	3
2.1.2 定义 .....	3
2.1.3 内涵及分类 .....	3
2.1.4 算法与程序的异同 .....	3
2.2 算法的描述及设计原则 .....	3
2.2.1 算法描述方法 .....	3
2.2.2 算法基本标准 .....	4
2.3 算法分析概论及有效算法 .....	4
2.3.1 概念 .....	4
2.3.2 复杂性分析 .....	6
2.4 算法设计方法概论 .....	7
2.4.1 概述 .....	7
2.4.2 算法设计的基本技术 .....	7
2.5 算法描述语言 .....	7
2.5.1 PDL 概述 .....	7
2.5.2 PDL 的优势 .....	8
2.5.3 PDL 书写及要求 .....	8

第三章 线性表 .....	13
3.1 线性表及其运算 .....	13
3.1.1 线性表的定义 .....	13
3.1.2 线性表的特征 .....	13
3.1.3 线性表的运算 .....	13
3.2 线性表的储存表示 .....	13
第四章 栈和队列 .....	17
4.1 栈及其运算 .....	17
4.1.1 绪论 .....	17
4.1.2 栈的基本定义 .....	17
4.1.3 与线性表的关系 .....	17
4.1.4 栈的运算 .....	17
4.1.5 栈的存储与运算的实现 .....	17
4.1.6 多栈共存问题 .....	17
4.2 栈的应用 .....	17
4.3 队列及其运算 .....	17
4.4 受限的栈及队列（了解） .....	18
第五章 串 .....	19
5.1 串及其运算 .....	19
5.2 串的模式匹配 .....	20
5.3 例题 .....	21
第六章 数组和广义表 .....	25
第七章 树形结构 .....	26
7.1 树的基本定义和运算 .....	26
7.1.1 树形结构的概述 .....	26
7.1.2 树的基本定义 .....	26
7.1.3 有序树的基本定义 .....	26
7.1.4 森林（树林） .....	27

7.2	二叉树 .....	27
7.3	遍历二叉树 .....	27
7.3.1	概念 .....	27
7.3.2	二叉树的遍历次序 (order) .....	27
7.4	树、森林与二叉树的转换 .....	28
7.4.1	概述 .....	28
7.4.2	树转为二叉树 .....	28
7.4.3	二叉树还原为树 .....	29
7.4.4	森林转为二叉树 .....	29
7.4.5	二叉树转为森林 .....	30
7.4.6	树的遍历 .....	30
7.4.7	森林的遍历 .....	31
7.5	线索树 .....	32
7.6	树形结构的应用 .....	32
	<b>第八章 latex 部分用法简介 .....</b>	<b>33</b>
8.1	用 latex 需要安装什么 .....	33
8.1.1	texlive 下载安装 .....	33
8.1.2	安装 IDE .....	34
8.2	常用的一些东西 .....	35
8.2.1	国际三线表格 .....	35
8.2.2	换页表格 .....	36
8.2.3	字体 .....	37
8.2.4	公式 .....	37
8.2.5	左边大括号 .....	38
8.2.6	复杂公式 .....	38
8.2.7	等号对齐站 .....	38
8.2.8	矩阵乘法 .....	38
8.2.9	图, 并列排 .....	38
8.2.10	附页代码 .....	39
8.2.11	参考文献 .....	40
	<b>参考文献 .....</b>	<b>41</b>

附录 .....	41
致谢 .....	42
论文（设计）成绩 .....	43

# 第一章 数据结构绪论

其实我推荐绪论写在正文里!! 作为第一章

这里是绪论，也可以说是引言，在 LZUThesis2020.clc 里面改，引言写什么呢，先凑字数，

是真的在打广告啊，嗯，做兰大毕业论文 LaTex 模板时，顺便介绍一下我写的软件：i 兰大易班，兰大专属的 app，可以看课表，充值校园卡（微信、支付宝都可以），查成绩（可以算绩点），还可以……，好多好多好多

注意啊，段落在 latex 里面是要空一行的，不要简单一个回车

## 1.1 数据结构的基本概念及研究内容

数据元素：具有完整确定意义的描述现实的某一个客观实体的一个最小数据集。数据元素类似原子，可以再分，每一项被称作数据项

数据对象：具有相同属性的数据元素的集合

数据结构：给定数据对象及其上面定义的操作所共同构成的一个系统一个信息处理模型

主要研究的三个方面：

### 1. 数据的逻辑结构

- 逻辑关系

在自然形态下，数据元素之间的一种关系

- 逻辑结构

数据之间所有关系的一个集合数学表示： $B=(K,R)$ ，其中：K：数据上的有穷集合 R：K 上关系的有穷集合，其中每个关系 r 都是从 K 到 K 的关系

- 分类

– 线性

    一对一，单对单

– 非线性

    \* 树形结构

        唯一一个直接前驱，多个直接后继

    \* 图结构

## 2. 数据的存储结构

- 存储关系

存储关系的数学内涵：须要建立数据对象 (K) 到存储区域 (M) 的映射关系 (S):

S:KM

即  $k \in K$ , 都有唯一的  $Z \in M$ , 使得  $S(k)=Z$ ,  $Z$  为  $k$  结点所占存储空间的始单元。

- 存储结构

- 顺序结构

按照连续地址空间的顺序依次的存放数据

- 链接结构

存储密度相比顺序结构下降

- 索引结构

- 散列结构

根据节点的值，通过一定的函数关系来确定数据元素的存储地址

## 3. 数据的运算关系定义在逻辑结构上，在存储结构上实施，即：

- 抽象层面

- 逻辑关系

- 需求

- 实现层面

- 存储关系

- 运算关系

- 评价层面

三个层次五个要素

## 1.2 数据结构的选择与评价

评价标准：

- 时间需要量时间效率
- 存储需要量存储效率

## 第二章 算法

### 2.1 算法的定义

#### 2.1.1 概述

算法是一个问题的具体解决方案

#### 2.1.2 定义

算法解决某一个问题的指令的有限集合

基本特征：

- 有穷性
- 确定性
- 可行性
- 输入
- 输出

#### 2.1.3 内涵及分类

内涵体现在过程上

- 一般过程
- 函数过程

强调结果

#### 2.1.4 算法与程序的异同

- 程序不是算法
- 程序不是算法程序不一定满足有穷性

### 2.2 算法的描述及设计原则

#### 2.2.1 算法描述方法

- 计算机程序设计语言  
优点与缺陷：设计出 = 实现出
- 自然语言缺陷：雍长二义性

- PDL
- 流程图

### 2.2.2 算法基本标准

- 正确性
- 易读性
- 健壮性
- 高效性

## 2.3 算法分析概论及有效算法

### 2.3.1 概念

一般不需要知道精确的时间消耗，需要知道时间消耗的增长率大体在什么范围。

算法复杂性的阶：算法比较主要比较阶

时间复杂性（时间渐进复杂性）：利用某算法处理一个问题规模为  $n$  的输入所需要的时间，记为  $T(n)$

空间复杂性：利用某算法处理一个问题规模为  $n$  的输入所需要的存储空间，记为  $S(n)$ 。

阶：对一个正常数  $C$ ，一个算法在时间  $(n^2)$  内能处理规模为  $n$  的输入，则称此算法的时间复杂度是  $(n^2)$ ，读作“ $n^2$  阶”，即该算法的时间复杂度与  $n^2$  是同阶的。

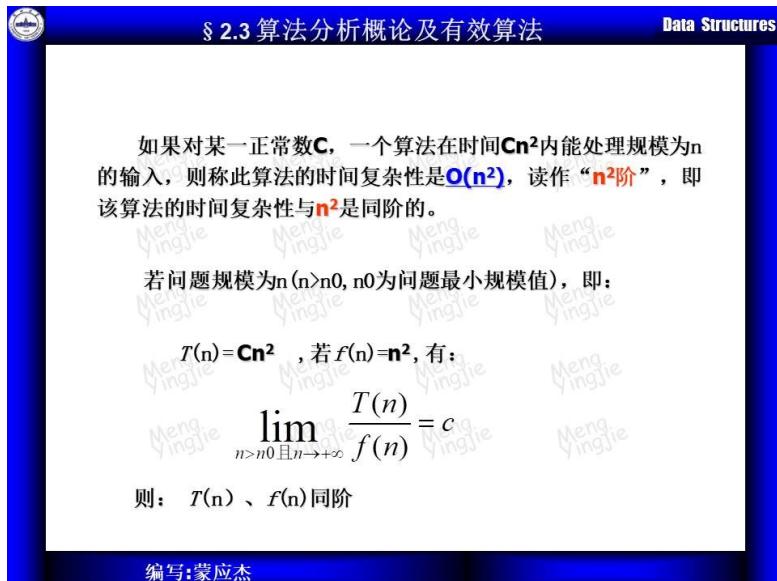


图 2.1 启动图形化安装界面

若一个算法时间复杂度为  $O(2n)$ , 称其需要指数时间; 若是  $O(nk)$ , 称其为多项式时间。当 n 非常大时两个时间差异非常大。

以多项式时间为界限的算法称为有效算法。如果一个问题不存在以多项式时间为界限的算法, 称为难解的 (难解性问题)

时间 复杂度	多项式时间算法与指数时间算法的运行时间对比					
	n=10	n=20	n=30	n=40	n=50	n=60
A算法	$10^{-7}$	$2*10^{-7}$	$3*10^{-7}$	$4*10^{-7}$	$5*10^{-7}$	$6*10^{-7}$
	$n^2$	$10^{-6}$	$4*10^{-6}$	$9*10^{-6}$	$1.6*10^{-5}$	$2.5*10^{-5}$
	$n^3$	$10^{-5}$	$8*10^{-5}$	$2.7*10^{-4}$	$6.4*10^{-4}$	$1.25*10^3$
	$n^5$	0.001	0.032	0.243	1.02	3.12
B算法	$2^n$	$10^{-5}$	0.001	10.74	3.05小时	130.3天
	$3^n$	$5.9*10^{-4}$	34.8	23.7天	3855年	$2*10^6$ 世纪
$n$ 为数据规模, 在每秒运行速度为1亿次的计算机上的运行时间 (单位:秒)						

编写:蒙应杰

图 2.2 启动图形化安装界面

当 n 非常大时两个时间差异非常大。

**§ 2.3 算法分析概论及有效算法**

**Data Structures**

**提高计算机的速度对单位时间内可解的问题规模的影响**

时间复杂度	单位时间内可解的问题规模		
	现在的计算机	快100倍的计算机	快1000倍的计算机
<b>A算法</b>	$n$	$N_1$	$100N_1$
	$n^2$	$N_2$	$10N_2$
	$n^3$	$N_3$	$4.64N_3$
	$n^5$	$N_4$	$3.98N_4$
	$2^n$	$N_5$	$N_5 + 6.64$
<b>B算法</b>	$3^n$	$N_6$	$N_6 + 4.19$
			$N_6 + 6.29$

从此表可看出，把计算机速度提高**1000**倍，**A算法**在单位时间内可解问题的规模是原来的**10**倍，而对于**B算法**，问题规模只增加**9.97 (≈10)**个。

编写:蒙应杰

图 2.3 启动图形化安装界面

对于时间复杂度，会关注时间复杂度的 最坏情况和时间复杂度的平均情况。

### 2.3.2 复杂性分析

不需要知道精确的数值，只需要知道他一个规模

- 时间复杂性的分析

1. 根据问题的特点合理选择一种或几种操作作为整个算法的“标准操作”（假设循环执行次数最多，循环就是标准操作）
2. 确定每个算法在给定输入下总共执行了多少次标准操作，并根据次数推导求出时间函数
3. 确定该函数的阶

- 空间复杂性的分析

1. 根据问题的特点合理选择一种或几种操作作为整个算法的“标准操作”（假设循环执行次数最多，循环就是标准操作）
2. 确定每个算法在给定输入下总共执行了多少次标准操作，并根据次数推导求出时间函数
3. 确定该函数的阶

要求各个算法的存储结构一样的前提下，关注算法所需的附加存储空间复杂性

## 2.4 算法设计方法概论

### 2.4.1 概述

- 用科学的方法进行算法设计
- 最常用方法：自顶向下逐步求精

顶层：问题的总和、抽象全貌

底层：问题的具体化、展开、细节

求精的方法：

1. 分而治之

将问题划分为一些不相交的部分，依次解决

2. 作出有限进展

采用一个朝解的方向得到有限进展的方法，反复应用，逐渐逼近

3. 情况分析

对问题各种情况给予分析，选择适当方案。

- 健壮性

- 高效性

### 2.4.2 算法设计的基本技术

以下内容不做要求

- 穷举法
- 分治法
- 回溯法
- 分支界限法
- 动态规划法
- 贪心法

## 2.5 算法描述语言

### 2.5.1 PDL 概述

Program、Design、Language

即伪码语言，主要用来书写软件设计的规约，是基于我们自然语言与具体的成设计语言之间的一种语言。这是一种保留计算机、程序设计、语言的基本框架和描述形式，并去掉一些特异性和直接性的要求，再结合自然语言所形成的一种用于描述算法处理的逻辑语言。

### 2.5.2 PDL 的优势

- 表达能力强，具有关键字的固定语法。提供了特定的结构化控制结构
- 引入了自然语言的一些习惯，结构比较清晰，简单易读
- 容易转化为任何一种程序设计语言代码（可由 PDL 生成程序代码）

### 2.5.3 PDL 书写及要求

#### 1. 算法的框架

- 一般过程的书写框架

```
1 PROC 过程名 (I/O参数) ;  
2 BEGIN  
3     语句组  
4 END;
```

- 函数过程的书写框架

```
1 FUNC 函数名 (I/O参数) : 类型名 ;  
2 BEGIN  
3     语句组  
4 END;
```

#### 2. 词的定义及说明

标识符：按照一定的规则形成的具有特定含义的一个词

- 过程名：调用前需定义
- 常量名、变量名：使用前需说明

例如 VAR i,j,k:integer

常见的数据类型名写法及表示：

- 整数型：integer
- 实数型：real
- 布尔型：boolean
- 字符型（单字符）：char
- 子介型（用于表达范围）：下界..上界，例如 40..90(表示 40-90), 'A'..'G'
- 枚举类型：0,1,2,3 元素次序不能变
- 构造类型
  - \* 数组型：

1 ARRAY[下标类型] OF 成分类型

例如：

1 A: ARRAY[1..20] OF integer

1 B: ARRAY[1..20, -10..20] OF real // (B是点集,  
二维数组)

\* 记录型：

1 RECORD

2 域标识符1：类型1

3 ...

4 域标识符n：类型n

5 END

例如：

1 A=RECORD

2 Name:ARRAY[1…8] OF char;

3 Sex:0..1

4 Age:integer;

5 END

- 指针类型：

1 类型名

例如：

1 TYPE A=integer; //指针类型

1 VAR B: integer; //指针变量

### 3. 基本语序

- 赋值语句：

1 变量名 表达式

- 流程图

- 条件语句

- 形式一

1 if 条件 then 语句组

- 形式二

```
1 if 条件 then 语句组1  
2           else 语句组2
```

- 循环语句

- 当型 (while)

```
1 WHILE 条件 DO  
2           语句组；
```

- 直到型 (repeat)

```
1 REPEAT  
2           语句组；  
3 UNTIL 条件
```

- 从到型 (for-to)

- \* 默认步长为 1:

```
1 FOR 变量 初值 TO 终值 DO  
2           语句组；
```

- \* 自定义步长:

```
1 FOR 变量 初值 TO 终值 STEP 步长值 DO  
2           语句组；
```

- \* 倒数:

```
1 FOR 变量 初值 DOWNTO 终值 DO  
2           语句组；
```

- 输入语句:

```
1 read(变量名表);
```

例:

```
1 read(x,y,z);
```

- 输出语句

```
1 write(变量名表);
```

## 4. 拓展语序

- 情况语句

```
1 CASE  
2     条件 1: 语句组 1;  
3     条件 2: 语句组 2;  
4     .....  
5     条件 n: 语句组 n;  
6     [ELSE 语句组 n+1]  
7 ENDCASE
```

- 一般过程调用语句:

```
1 Call 过程名;
```

- 函数过程调用: 通过在表达式中引用函数名完成, 即被引用函数名出现在表达式中。
- 出错提示语句:

```
1 error(错误信息);
```

- 终结语句

```
1 Exit \\ 算法转向正常结束
```

```
1 Return \\ 算法转向正常结束, 携带值离开
```

```
1 Abort \\ 中途废止 (中止)
```

- 复合语句

```
1 [ 简单语句 1;  
2     简单语句 2;  
3     .....  
4     简单语句 n; ]
```

或用 Begin End 代替括号

- 动态符号

– 储存单元的引用:

```
1 指针变量名
```

例:

```
1 x
```

– 动态空间分配:

1 New(P)

– 动态空间回收:

1 Dispose(P)

– 空地址的表示:

1 Nil

## 第三章 线性表

### 3.1 线性表及其运算

#### 3.1.1 线性表的定义

线性表的定义：一个线性表是  $n_0$  个数据元素  $a_1, a_2, \dots, a_n$  的有限序列，序列中除第一及最后一个元素以外，每个元素有且只有一个直接前驱和直接后继。

简称表，可表示为：  $A = (a_1, a_2, \dots, a_n)$

```
1 ai:datatype //表示  $a_i$  项可以是任何类型
```

#### 3.1.2 线性表的特征

- 有限的。线性表的表长：线性表元素的个数。控标的长度定义为 0
- 元素呈线性关系。元素的位置只取决于他们自己的逻辑顺序。

#### 3.1.3 线性表的运算

- 确定线性表的长度  $n$
- 存取线性表的第  $i$  个数据元素，检验或改变某个数据项的值
- 在第  $i-1$  个和第  $i$  个数据元素之间插入一个新的数据元素。约定插入的元素是第  $i$  个元素的直接前驱
- 删去第  $i$  个元素
- 将两个或两个以上的线性表合并成一个线性表
- 将一个线性表拆分成两个或两个以上的线性表
- 重新复制一个线性表
- 对线性表中的数据元素依据某一种规则进行重组

### 3.2 线性表的储存表示

- 线性表的向量表示
  - 存储方法顺序地分配存储单元，且每个数据元素占据相同大小的存储空间（顺序且等长）
  - 数据访问 TODO
  - 向量存储结构特性

- \* 储存分配呈线性结构
- \* 属于随机存储结构（访问一个元素的代价与元素位置无关，即访问任何一个元素（找地址）的运算量一样，一维数组，通过下标变量来访问（数组构造的本质即算公式））
- 向量存储结构的形式化表示可用一个一维数组表示，由于数组属于静态结构，其空间规模须事先定义（1..max），要有个计数器记录数组长度（空间规模）

表述形式：

```

1 TYPE SQLIST:ARRAY [1..max] OF datatype;
          ## 内容
2 VAR n:0..max;
          ## 计数器（记录
          数组长度）

1 TYPE SQLIST= RECORD
2
          data:ARRAY
          [1..
          max
          ]
          OF
          datatype
          ;
          ## 内容 n:0..max; ## 计数器（记录数组长度）
4      END;

```

- 插入
- 删除
- 小结
- 线性表的链表表示
  - 单链表表示

```

1 TYPE pointer = node;
2           node=RECORD
3                   data: datatype;
4                   next: pointer;
5           END;
6           link=pointer;

```

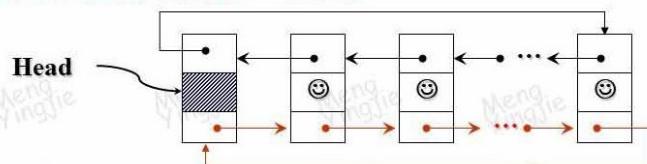
- 带表头的单链表表示
- 带表头结点的循环单链表表示
- 带表头结点的双向循环链表表示

```

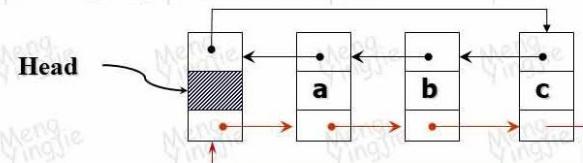
1 TYPE pointer = node;
2           node=RECORD
3                   Left: pointer;
4                   data: datatype;
5
6                   right: pointer;
7           END;
8           dblink=pointer;

```

### 带表头的双向循环链表的一般表示:



例1，线性表A=(a,b,c)的存储示意图：



例2，线性表A=() 的存储示意图：

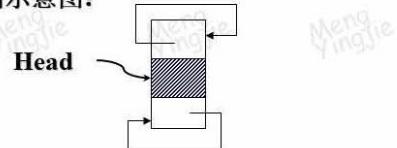
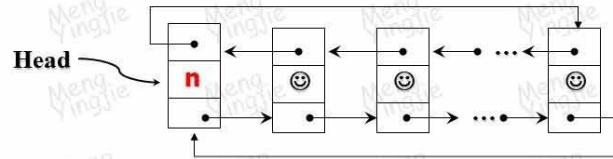


图 3.1 启动图形化安装界面

实际使用中双向链表的一些其它变化：

例1. 表头结点的数据项的利用：



例2. 指针数据项的变化：

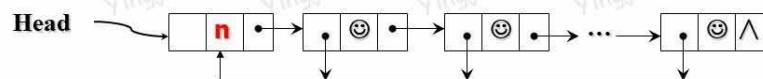


图 3.2 启动图形化安装界面

- 线性表的应用

## 第四章 栈和队列

### 4.1 栈及其运算

栈是工具性数据结构

#### 4.1.1 绪论

#### 4.1.2 栈的基本定义

栈是一个下限为常数，上限可变化的向量（或者反之）。有时称为堆栈或堆阵。

后进先出表（LIFO）

可变化一端称为栈顶，不变化的一端称为栈底

#### 4.1.3 与线性表的关系

- 相同点

- 逻辑关系都是线性关系

- 不同点

- 线性表可以从任意位置增删，栈的增删只能从表尾进行

- 栈的运算是线性表的一个子集，且这个子集还需加以约束

#### 4.1.4 栈的运算

- 入栈，PUSH(S)，完成往栈中加入元素的过程，即插入操作，也称为压栈
- 出栈，POP(S)，完成从栈中取出元素的过程，即删除操作，也称为弹出

#### 4.1.5 栈的存储与运算的实现

#### 4.1.6 多栈共存问题

### 4.2 栈的应用

### 4.3 队列及其运算

- 绪论分时和并行，主要处理对稀缺资源的争夺
- 队列的定义队列是一个下限和上限只能增加而不能减少（下限和上限的指针只能往一个方向移动）的向量（或者反之）

### 先进先出表 (FIFO)

- 队列与线性表
  - 相同点
    - \* 逻辑关系都是线性关系
  - 不同点
    - \* 线性表可以从任意位置增删，队列的只能一端插入一端删除
    - \* 队列的运算是线性表的一个子集，且这个子集还需加以约束
- 队列的运算
  - 出队
  - 入队
- 队列的存储与运算的实现

## 4.4 受限的栈及队列（了解）

- 双端队列

双端队列是一种所有的插入和删除都限制在表的两端进行的线性表

- 双栈

双栈是一种加限制的双端队列，即从哪端进就只能从哪端出，就像是两个底部相连的栈

- 超队列

超队列是一种删除受限制的双端队列，删除限制在一端，插入可以在两端

- 超栈

超栈是一种插入受限制的双端队列，插入限制在一端，删除可以在两端

## 第五章 串

### 5.1 串及其运算

- 概述

字符串是线性表模型数据元素实例化的体现

全称字符串

早期：作为输入输出的常量和提示

热点：中文信息处理

- 串的定义

一个由零个或多个字符组成的有穷序列称为串

简记为  $A = 'a_1 \ a_2 \ a_3 \dots \ a_n'$

串的长度：串中所含的字符个数

空串：串长为零的串，记作

非空串

空白串：空白字符组成的串，记作“”

串的相等：串长也相等，对应位置上的字符也一样

字串/主串：一个串中任意个连续字符组成的子序列称为该串的子串，该串成为它的所有子串的主串。不一定是一个真子集

- 串的运算

- 赋值

`assign(S,chars)` 将字符串常量 `chars` 赋给字符串变量 `S`

- 连接

`concatenation(S,T)` 将字符串 `S` 和 `T` 联接在一起

- 取子串

`substring(S,m,n)` 从第 `m` 个位置开始取 `n` 个连续字符（通常）/从第 `m` 个位置开始到第 `n` 个位置

- 求子串序号

`strindex(S,T,i)` 确定串 `T` 在 `S` 中第一次出现的位置 `i`

- 串的插入

`strinsert(S,T,i)`

- 串的删除
  - strdelete(S,m,n)
- 串的复制
  - copy(S,T)
- 串的置换
  - replace(A,B,C) (找到 A 中的 B 用 C 来替换)
  - replace(S,m,n,T) (把 S 字符串的 (m 到 n/从 m 开始的 n 个字符) 用 T 来替换)
- 串的存储
  - 顺序储存
    - 一般用向量来表示，通常称为字符串数组、字符串变量、字符串常量
    - \* 压缩模式
      - 按字节存储数据 (C 语言)
    - \* 非压缩格式
      - 运算器以字为单位运算，存储器以字节为单位，为了避免转换浪费的时间，存储器以字为单位存储一个字符，变成非压缩模式（只存一个字母需要一个字的空间，比较浪费）
  - 索引储存
    - 多用在对于多个字符串常量或者变量的组织
  - 链接储存
    - 理论上研究的多，实际上用的少，不易于实现

## 5.2 串的模式匹配

- 概述

在模式分类或者问题回答系统等方面，将输入模式与样本模式进行匹配的过程啊，我们就把它称为模式匹配

通常把 S 称为目标 (或正文)，把 P 称为模式，把从目标 S 中查找模式 P 的过程称为模式匹配

串的匹配是模式匹配的实例化
- 匹配的朴素算法 (Brute-Force 算法)

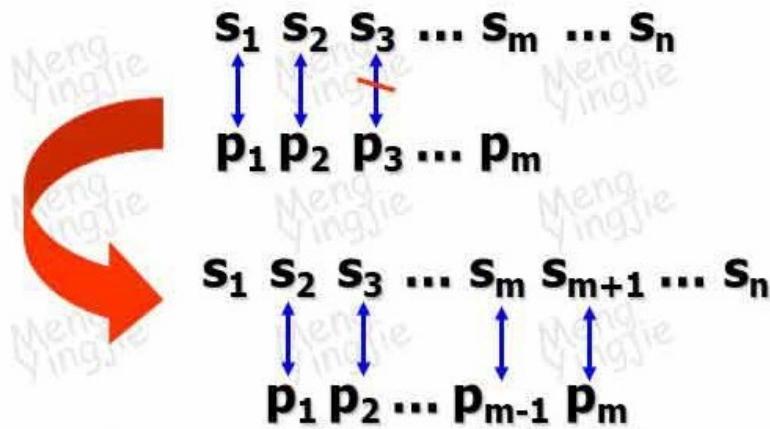


图 5.1 启动图形化安装界面

处理基本思想：对正文顺序搜索

1. 逐次比较（对应字符依次比较）
2. 发现不匹配时，将 P 相对于 S 右移一位
3. 重复上述过程，直到成功或扫描完 S

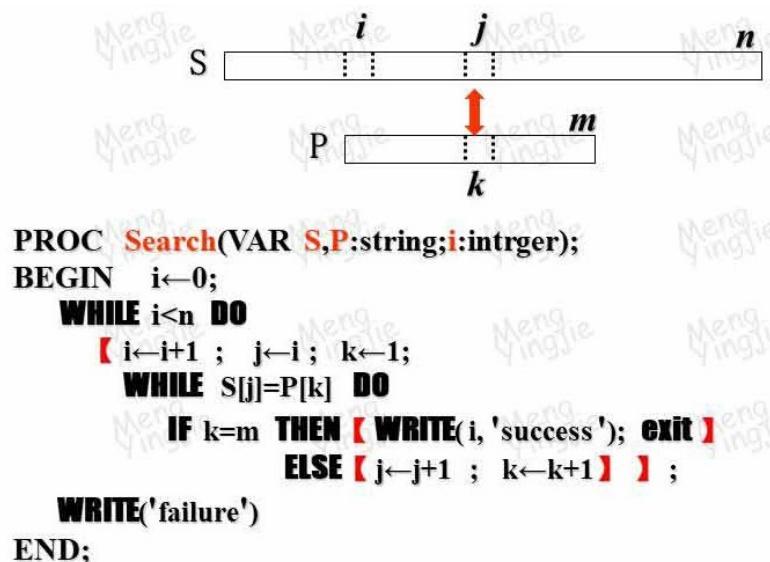


图 5.2 启动图形化安装界面

优化算法：KMP 算法

### 5.3 例题

2019 年兰州大学开源社区纳新面试题中有一题关于字符串的运算，现将我的解决方案附于后文。

给出一个文本文件，其中每个单词不包括空格及跨行，单词由字符序列构成且不区分大小写，完成以下功能：统计给定单词在文本文件中出现的总次数。

注：允许使用 string.h

```
1 #include<stdio.h>
2 #include<string.h>
3
4 char word[100];
5 char text[10000];
6 int count = 0;
7
8 char trans(char c){
9     if(c >= 'A' && c <= 'Z'){
10         c = c + 'a' - 'A';
11     }
12     return c;
13 }
14
15 void getword() {
16     char c;
17     int i = 1;
18     while (1) {
19         c = getchar();
20         if (c == '\n') break;
21         word[i] = trans(c);
22         i++;
23     }
24     word[0] = i - 1; //第0位存放单词长度
25     return;
26 }
27
28 void gettext() {
29     char c;
30     int i = 1;
31     while (1) {
```

```
32         c = getchar();
33         if (c == EOF) break;
34         text[i] = trans(c);
35         i++;
36     }
37     text[0] = i - 1; //第 0 位存放文本长度
38     return ;
39 }
40
41 void check() {
42
43     int i = 0, j = 0;
44     int flag = 1;
45     while (1) {
46         i++;
47         j++;
48         if (i == text[0]) break;
49         if (text[i] == '\n' || text[i] == ' ') {
50             j = 0;
51             flag = 1;
52             continue;
53         }
54         if (flag == 0) {
55             continue;
56         }
57         if (text[i] != word[j])
58             flag = 0;
59         if (j == word[0]) {
60             if (flag == 1) {
61                 if (text[i + 1] == '\n' || text[i + 1] == ' ')
62                     count++;
63             }
64             flag = 1;
```

```
65 }
66 }
67 }
68
69 int main()
70 {
71     freopen( "C:\\Users\\Kente\\Desktop\\123\\text.txt" , "r"
72         , stdin ); //假设给出的文本文件第一排是给定单词，第二
73         //排及之后是要统计的文本
74     getword();
75     gettext();
76     check();
77     printf( "%d" , count );
78     return 0;
79 }
80 /*text.txt
81 ABC
82 Abc abC dasasdsabc
83 dasdasdasdasd dasdasdasdasd
84 abc ab
85 */
```

## 第六章 数组和广义表

## 第七章 树形结构

### 7.1 树的基本定义和运算

#### 7.1.1 树形结构的概述

- 与线性结构的差异  
直接后继可以是多个
- 树形结构的地位  
树是计算机中最重要的非线性结构
- 讨论内容
  - 逻辑角度
  - 存储方式
  - 应用
  - 模型
    - \* 树
    - \* 二叉树

#### 7.1.2 树的基本定义

树  $T$ , 是满足如下性质的有限个节点组成的非空集合

- $T$  中有且仅有一个称为根的结点
- 除根结点之外, 其余结点分成  $m(m>0)$  个不相交的集合  $T_1, T_2, \dots, T_m$ , 其中每个  $T_i$  都是树, 而且都称为  $T$  的子树。

注意

- 是递归定义
- 定义中对于子树的个数及次序没有进行任何约束
- 与图论中的树不同, 是有向图中的根树的特例。隐含了向下的方向性

#### 7.1.3 有序树的基本定义

在树  $T$  中如果子树  $T_1, T_2, \dots, T_n$  的相对次序是重要的 (即有序的), 则称  $T$  为有向有序树, 简称有序树。

默认树是有序树

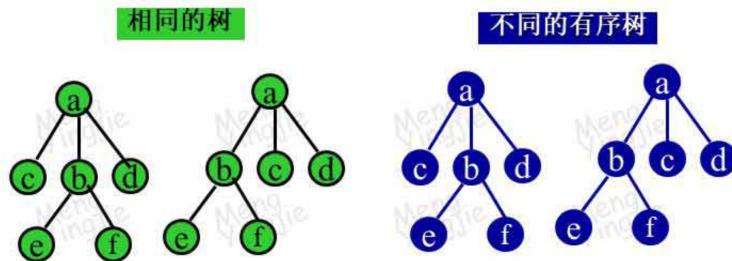


图 7.1 启动图形化安装界面

#### 7.1.4 森林（树林）

森林是零棵或多棵不相交的树的集合（通常是有序集合）

## 7.2 二叉树

- 定义
- 特性
- 二叉树的存储表示

## 7.3 遍历二叉树

### 7.3.1 概念

定义：对于给定的数据结构，系统地访问该结构中的每个结点，且每个结点仅被访问一次的操作过程称遍历

系统地：按照一定的规律、次序

访问：对元素进行的某种操作

遍历本质上是一个运算过程（序列）

对于给定的二叉树，系统地访问该结构中的每个结点，且每个结点仅被访问一次的操作过程称遍历

### 7.3.2 二叉树的遍历次序（order）

- 层次策略
  - 自上而下
    1. 从左到右
    2. 从右到左
  - 自下而上

- 深度策略遍历针对结构，是一个过程；访问针对一个点，是一个动作。

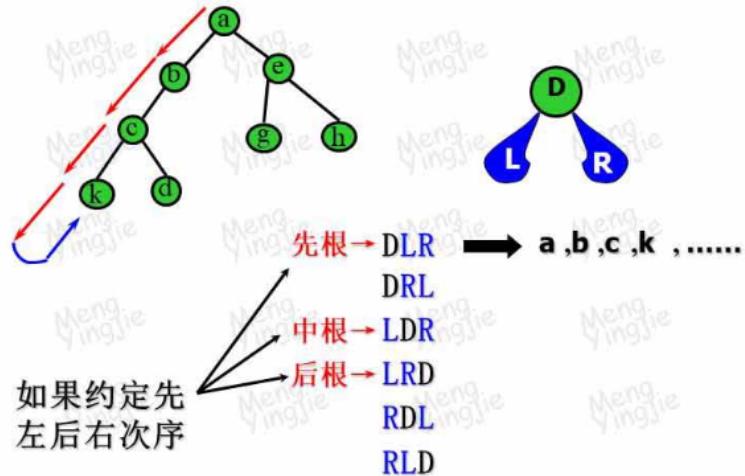


图 7.2 启动图形化安装界面

- 先根（前序）DLR

先拿到根，然后从左子树走，又是二叉树，则访问了左子树的根节点，再往下走，直到访问到 k，左子树为空，回到 c，访问右子树。沿着纵深往下搜索，只要有子树就往下搜索，先不管兄弟。

- ### - 中根 LDR

- 后根（后序）LRD

## 7.4 树、森林与二叉树的转换

#### 7.4.1 概述

转换原因

- 二叉树有许多优良特性，而树没有
  - $n$  叉树中二叉树的空间利用率最高

### 7.4.2 树转为二叉树

这里约定树为有序树，子树的次序相对固定

转换规则：

1. 加线（亲兄弟之间加上虚连线）
  2. 抹线（除了最左子节点之外，抹掉所有节点与其余子节点的连线）
  3. 调整（调成二叉树的样子，原有连线在左，新加连线在右）

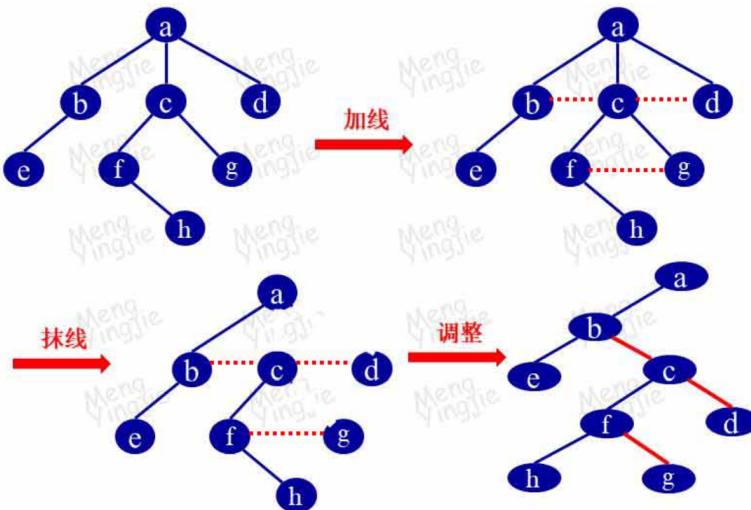


图 7.3 启动图形化安装界面

特点：

- 原有兄弟成为右结点及右结点的右列结点
- 根子树仅有左节点

#### 7.4.3 二叉树还原为树

转换规则：

1. 加线（如果有节点 X 是父结点的左子树，那么它的右子树，右子树的右子树... 都与 X 的父节点加线）
2. 抹线（所有节点抹掉与右子树的连线）
3. 调整（让节点按层次分布）

#### 7.4.4 森林转为二叉树

两种方法：

- 方法一：树转换  
将每棵树转为二叉树
- 方法二：二叉树连接

依据转换得到的二叉树的次序，将后一棵作为前一棵根节点的右子树，因为树转二叉树的根节点无右子树

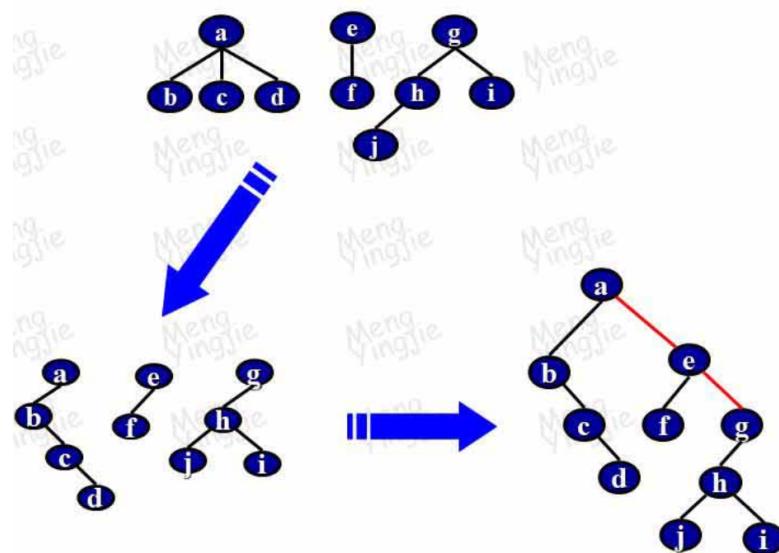


图 7.4 启动图形化安装界面

#### 7.4.5 二叉树转为森林

转换规则：

1. 抹线（沿着根节点，抹掉其右子树与根节点的连线，不断往右下搜索）
2. 还原（把每棵二叉树还原树）

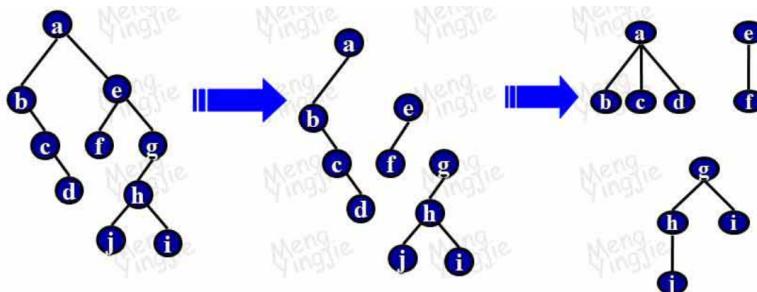


图 7.5 启动图形化安装界面

#### 7.4.6 树的遍历

- 方法一：树转换

将每棵树转为二叉树后再处理，处理完成后还原

- 方法二：直接处理

遍历次序

– 层次策略

\* 自上而下

1. 从左到右
  2. 从右到左
- \* 自下而上
- 深度策略
- \* 先根（前序）DLR
    1. 访问树的根节点
    2. 先根次序下依次遍历树的根节点的每个子树
  - \* 中根 LDR  
一般不讨论
  - \* 后根（后序）LRD
    1. 后根次序下依次遍历树的根节点的每个子树
    2. 访问树的根节点

#### 7.4.7 森林的遍历

- 方法一：树转换

将每棵树转为二叉树后再处理，处理完成后还原

- 方法二：直接处理

遍历次序

- 层次策略

\* 自上而下

1. 从左到右
2. 从右到左

\* 自下而上

- 深度策略

\* 先根（前序）DLR

1. 访问第一棵树的根节点
2. 先根次序下依次遍历第一棵树的根节点的每个子树
3. 先根次序下依次遍历其余的树

\* 中根 LDR

一般不讨论

\* 后根（后序）LRD

1. 后根次序下依次遍历第一棵树的根节点的每个子树
2. 访问第一棵树的根节点
3. 后根次序下依次遍历其余的树

## 7.5 线索树

## 7.6 树形结构的应用

## 第八章 latex 部分用法简介

注意啊，看这个教程，template.pdf 配合 template.tex 一起看，才能学习 latex 怎么用的

网页跳转怎么用？图片插入怎么用？图片横着两个并排站呢？代码怎么插入？表格听说挺复杂？公式听说也挺难的

哈哈哈，你说你还不知道什么是 LaTeX，你去分不清 XeLaTeX、pdfLaTeX，百度一下竟然还让我安装 TexLive，这也就算了，甚至还有人说 vscode? sublime text3? texstudio? Texmaker? 我只是想写个论文排版方便一些，你要干嘛？

上面这些问题，后面都会一点点介绍

### 8.1 用 latex 需要安装什么

需要安装 texlive，外加一个 IDE

#### 8.1.1 texlive 下载安装

最近可能出了 2020 了，可以用兰大的镜像下载应该在用校园网时快一些，额，你还是用清华的镜像吧，我刚才找了一下，兰大镜像这会儿竟然挂了。。。

下载地址<sup>1</sup>：点下面的字跳转浏览器下载了，方便吧

- TexLive2019 Windows 版
- TexLive2019 Mac 版

上面的文件直接双击安装一路 next 就行，但是 texlive 这是个啥？

用过 python 吧，texlive 就相当于你下载的 python 安装包，但是你总不能在终端里写代码吧，一般用 pycharm，这个就是 IDE，所以你需要再安装一个 IDE。

为什么没 linux 版？用的人不多，真心不想给。。。其实安装文件就是 windows 的那个版本

#### linux 系统图形界面安装 texlive

1. 安装 perl 组件：sudo apt-get install perl-tk
2. 加载该 ISO 文件：sudo mount -o loop texlive2019.iso /mnt （换掉文件路径即可）<sup>2</sup>

<sup>1</sup>这个地址会自动更新比如 2020 版出了以后你下载的就是 2020 了

<sup>2</sup>注意：使用该命令会出现错误提示，mount: /dev/loop1 is write-protected, mounting read-

3. 启动图形化安装界面: cd /mnt & sudo ./install-tl -gui

注意倒数第二项, 改成是, 创建符号链接, 下面那个图是网上随便找的, 都差不多

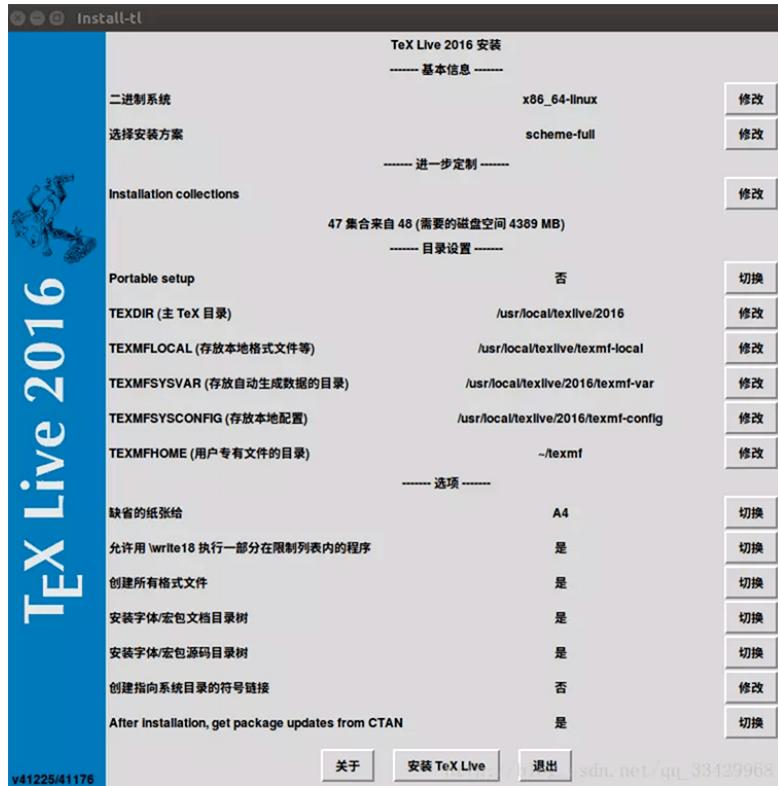


图 8.1 启动图形化安装界面

### 8.1.2 安装 IDE

在这之前, 请测试 texlive 是否安装成功!!! 在命令行输入 tex, 显示类似如下结果, 注意必须包含 “TeX Live 2020” 字样

```
1 This is TeX, Version 3.14159265 (TeX Live 2020) (
    preloaded format=tex)
2 **
```

如果确实安装了, 但是没有显示, 请根据各自系统自行百度配环境变量, 此处不再详细介绍

IDE 这个就是写论文的地方, 它会调用你刚才安装的 texlive, 具体用什么, 各有所爱, 我喜欢 sublime text3, 这个颜值是真的高, 而且体积小, 启动快, 可以预览公式什么的, 很多常用的代码可以自动提示补全, 但是这个需要安装插件

---

only. 不必管它

LaTexTools, pdf 需要安装其他的东西进行正反向跳转<sup>1</sup>, 小白的话就算了吧, 想折腾, 百度一下吧

另外一个我比较喜欢的是 vscode, 也需要安装插件, 很方便, 和 sublime text3 差不多, 但是用起来简单一些, 也自己百度去吧

另外几个是 Texmaker、或者 texstudio<sup>2</sup>, 这两个你点名字就跳转官网了, 这两个基本上是打开就可以用, 怎么用, 自己百度吧, 很多详细的图文教程

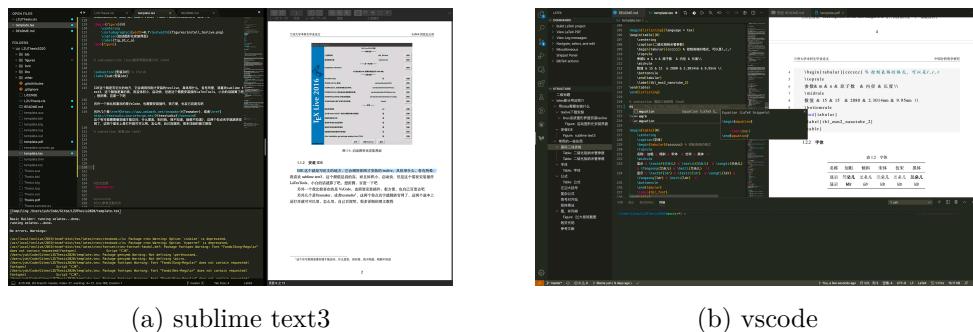


图 8.2 我用的 IDE

这两个 IDE 真的是特别好用, 不要用 TexMaker 或者 TexStudio 了, 连个自动提示都麻烦, 预览也没有, 最主要的是太丑了, 也不能换主题

## 8.2 常用的一些东西

用到相关的直接到这里复制, 然后修改就行

### 8.2.1 国际三线表格

表 8.1 二硫化钼纳米管参数

参数	m	n	换行一下	内径	长度
数值	15	15	2880	2.3014nm	9.95nm

<sup>1</sup>就是点 tex 文件某一行跳转到 pdf 对应的地方, 点击 pdf 跳转到 tex 对应的那一行, mac 上安装 skim, windows 安装 sumatra

<sup>2</sup>这个有可能需要翻墙才能访问, 什么意思, 别问我, 我不知道, 啥都不知道

这个注意,有多少列,后面就要有多少个 c<sup>1</sup>,这个 c 表示这一列居中 (center),靠左的话: l, 右: r;

那个 label 后面的名字自己取,但是不能有重复,是为了引用,比如这样,表格8.2,方程、图片也是这样引用的,好处是,中间加一个表格导致这个表格的序号变了也没事,你不用再去修改其他地方的引用

```

1 \begin{table}[H]
2     \centering
3     \caption{二硫化钼纳米管参数}
4     \begin{tabular}{cccccc} \% 控制表格的格式, 可以是 l, c, r
5         \toprule
6         参数&m &n & 原子数 & 内径 & 长度 \\
7         \midrule
8         数值 & 15 & 15 & 2880 & 2.3014nm & 9.95nm \\
9         \bottomrule
10    \end{tabular}
11    \label{tbl_mos2_nanotube_2}
12 \end{table}
```

### 8.2.2 换页表格

我是真的没想到有的人表格居然这么长,竟然能有三页。。。

表 8.2 二硫化钼纳米管参数

参数	m	n	原子数	内径	长度
数值	15	15	2880	2.3014nm	9.95nm
数值 1	15	15	2880	2.3014nm	9.95nm
数值 2	15	15	2880	2.3014nm	9.95nm
数值 3	15	15	2880	2.3014nm	9.95nm
数值 4	15	15	2880	2.3014nm	9.95nm
数值 5	15	15	2880	2.3014nm	9.95nm
数值 6	15	15	2880	2.3014nm	9.95nm
数值 7	15	15	2880	2.3014nm	9.95nm
数值 8	15	15	2880	2.3014nm	9.95nm

<sup>1</sup>否则会报错: Extra alignment tab has been changed to cr. 有什么报错百度一下一般就找到了

数值 9	15	15	2880	2.3014nm	9.95nm
数值 10	15	15	2880	2.3014nm	9.95nm
数值 11	15	15	2880	2.3014nm	9.95nm
数值 12	15	15	2880	2.3014nm	9.95nm
数值 13	15	15	2880	2.3014nm	9.95nm
数值 14	15	15	2880	2.3014nm	9.95nm
数值 15	15	15	2880	2.3014nm	9.95nm
数值 16	15	15	2880	2.3014nm	9.95nm
数值 17	15	15	2880	2.3014nm	9.95nm
数值 18	15	15	2880	2.3014nm	9.95nm
数值 19	15	15	2880	2.3014nm	9.95nm
数值 20	15	15	2880	2.3014nm	9.95nm

### 8.2.3 字体

表 8.3 字体

名称	加粗	倾斜	宋体	仿宋	黑体
显示	兰朵儿	兰朵儿	兰朵儿	兰朵儿	兰朵儿
显示	ldr	ldr	ldr	ldr	ldr

发现没，中文斜体没有效果的，你可以自定义，这个自己百度吧；而且加粗也 windows 系统上也是没有效果的，一般都改成了黑体（比如这个模板中成绩页等加粗的地方都是用的黑体），当然你也可以自定义。怎么做，百度吧

### 8.2.4 公式

所有的符号都要用美元符号包裹 \$，需要用到某一个但是不知道，直接百度，基本上都有

表 8.4 公式

名称	分数	下角标	上角标	矢量	根号	希腊字母	点乘	叉乘	矢量
显示	$\frac{1}{2}$	$O_2$	$a^2$	$\vec{AB}$	$\sqrt[3]{3}$	$\theta$	$\cdot$	$\times$	$\vec{a}$

但是有时候我们只是正文中想用  $MoS_2$ ，它竟然斜体，不想斜体，我写了个命令，这样用  $MoS_2$ ，正的吧，常用的命令可以自定义

### 8.2.5 左边大括号

$$\begin{cases} \vec{e}_1 = \frac{3a}{2}\vec{i} + \frac{\sqrt{3}a}{2}\vec{j} \\ \vec{e}_2 = \frac{3a}{2}\vec{i} - \frac{\sqrt{3}a}{2}\vec{j} \end{cases} \quad (8.1)$$

注意后面有个方程的编号,如果想取消,把上下的两个 *equation* 改成 *equation\**

$$\begin{cases} \vec{e}_1 = \frac{3a}{2}\vec{i} + \frac{\sqrt{3}a}{2}\vec{j} \\ \vec{e}_2 = \frac{3a}{2}\vec{i} - \frac{\sqrt{3}a}{2}\vec{j} \end{cases}$$

### 8.2.6 复杂公式

不会输出的符号,请百度,啥都有

$$\hat{H} = \frac{\epsilon}{2}\hat{\sigma}_z - \frac{\Delta}{2}\hat{\sigma}_x + \sum_k \omega_k \hat{b}_k^\dagger \hat{b}_k + \sum_k \frac{g_k}{2} \hat{\sigma}_z (\hat{b}_k + \hat{b}_k^\dagger) \quad (8.2)$$

### 8.2.7 等号对齐站

主要是用这个 aligned 放在了方程的环境里,等号前面 & 控制对齐,每一行后面双斜杠换行

$$\begin{aligned} \vec{CH} &= m \cdot \vec{e}_1 + n \cdot \vec{e}_2 \\ &= \frac{3(m+n)a}{2}\vec{i} + \frac{\sqrt{3}(m-n)a}{2}\vec{j} \end{aligned} \quad (8.3)$$

### 8.2.8 矩阵乘法

其实就是几个 array 组合

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (8.4)$$

### 8.2.9 图,并列排

这一句代表这个图片宽度为一行文本宽度的  $\frac{3}{10}$

```
1 width=0.3\textwidth
```



图 8.3 i 兰大易班截图

### 8.2.10 附页代码

可以在 LZUThesis.clc 里面修改代码格式

java 代码

```
1 System.out.print("i 兰大易班")
2 // 试一下中文注释
```

tex 代码

```
1 width=0.3\textwidth
2 % 注释
```

python 代码

```
1 print("i 兰大易班")
2 # 注释
```

matlab 代码有专门的库，但是没必要高亮太多，而且中文适配有问题，直接按照下面这个就可以

```
1 display("i 兰大易班")
2 % 注释
```

### 8.2.11 参考文献

这个，百度学术、谷歌学术等网站都可以导出 bibtex 格式的参考文献（知网不行，网上有个人写了个转换器，但是 windows 用不了，就不放了，尽量用谷歌学术把那个文献找出来吧），直接放在 bib/database.bib 文件里、知网需要用其他东西转换，但是我建议用 mendeley 这个软件管理文献，然后可以导出 bibtex 格式的，甚至可以直接复制引用，很方便 [?, ?, ?]。

有些人希望多个参考文献同时引用时用 [1-3] 而不是 [1,2,3]，所以我加了个包 cite。（2020-5-18）

具体怎么用可以百度，我这里告诉你什么可以用，但是具体的，建议百度，更靠谱一些。

有参考文献时，编译要经过 4 步，直接 XeLaTeX → BibTeX → XeLaTeX → XeLaTeX，不然很多问题，我用的 sublime text3，配合插件 LatexTool，直接快快捷键 ctrl - B，就可以自动完成 4 步了，很方便

## 附录

这里是附录页，附上你的程序或必要的相关知识

编译方式: XeLaTeX → BibTeX → XeLaTeX→XeLaTeX

## 致 谢

这里是致谢页，你可以在这里致谢你的舍友，老师，朋友，或者我。

### 论文（设计）成绩表

导师评价你人很好

建议成绩 80

指导教师（签字）



答辩小组意见

优秀

答辩委员会负责人（签字）

成绩 100

学院（盖章）

年 月 日